

Ejercicio 3: Comparación de Métodos Experimentales y Observacionales

JTRAIN2 vs JTRAIN3 - Capacitación Laboral

Contexto del Problema

Queremos evaluar el efecto de programas de **capacitación laboral** (*job training*) sobre los **ingresos reales** (*real earnings*). La pregunta de investigación es:

¿Aumenta la participación en capacitación laboral los ingresos de los trabajadores?

Dos enfoques para responder esta pregunta

1. JTRAIN2: Datos Experimentales

- **Diseño:** Experimento controlado aleatorizado (RCT - Randomized Controlled Trial)
- **Asignación al tratamiento:** Los investigadores asignan **aleatoriamente** a los individuos a:
 - Grupo de tratamiento: Reciben capacitación ($\text{train} = 1$)
 - Grupo de control: No reciben capacitación ($\text{train} = 0$)
- **Ventajas:**
 - Aleatorización garantiza que ambos grupos sean **comparables** en todas las características (observables y no observables)
 - Identificación causal clara: $E[Y_i(1) - Y_i(0)|X_i] = E[Y_i(1) - Y_i(0)]$
 - Alta **validez interna**
- **Desventajas:**
 - Costoso y difícil de implementar
 - Muestra específica puede limitar **validez externa**
 - Consideraciones éticas en algunos contextos

2. JTRAIN3: Datos Observacionales

- **Diseño:** Datos observacionales con **auto-selección**
- **Asignación al tratamiento:** Los individuos **eligen** participar o no
- **Ventajas:**
 - Refleja decisiones reales en el mercado laboral
 - Más barato y fácil de recolectar
 - Mayor **validez externa** (si la muestra es representativa)
- **Desventajas:**
 - **Sesgo de selección:** Quienes eligen participar pueden ser sistemáticamente diferentes
 - Grupos tratado y control no son comparables
 - Requiere supuestos fuertes para inferencia causal
 - Menor **validez interna**

Pregunta a: Fracción que recibió capacitación

Resultados

Dataset	Fracción tratada	Interpretación
JTRAIN2 (Experimental)	$\approx 36 \%$	Asignación aleatoria
JTRAIN3 (Observacional)	$\approx 11 \%$	Auto-selección

Cuadro 1: Proporción de participantes en capacitación

¿Por qué hay una diferencia tan grande?

La diferencia en las proporciones tratadas refleja **diferencias fundamentales** en el diseño de cada estudio:

1. Diseño experimental (JTRAIN2 - 36 %)

En el experimento:

- Los investigadores **controlan** quién recibe el tratamiento
- La proporción tratada (36 %) fue **elegida** por los investigadores para maximizar:
 - Poder estadístico (más observaciones en tratamiento)

- Restricciones presupuestarias (capacitar es costoso)
- Balance entre grupos
- Esta proporción **NO refleja** demanda real del mercado

Fórmula del poder estadístico: Para un test de diferencia de medias, el poder aumenta cuando la proporción tratada se acerca a 50 %:

$$\text{Poder} \propto \sqrt{\frac{1}{p(1-p)}} \quad (1)$$

donde p es la proporción tratada. El máximo se alcanza en $p = 0,5$.

2. Diseño observacional (JTRAIN3 - 11 %)

En los datos observacionales:

- La proporción tratada (11 %) refleja la **demanda real** de capacitación
- Factores que limitan la participación:
 - **Costos de oportunidad:** Participar requiere tiempo que podría usarse trabajando
 - **Barreras de información:** No todos conocen el programa
 - **Auto-selección positiva:** Solo quienes esperan beneficiarse participan
 - **Restricciones de elegibilidad:** Algunos no califican

3. Implicaciones de la diferencia

Esta diferencia en proporciones es **síntoma de un problema más profundo:**

Advertencia Crítica

La baja tasa de participación en JTRAIN3 (11 %) sugiere fuerte **auto-selección:**

- Quienes participan pueden ser los más **motivados**
- O los más **desesperados** (desempleados crónicos)
- O aquellos con **mayores retornos esperados**
- O aquellos con **menores costos de oportunidad**

Esto genera **sesgo de selección** que invalida comparaciones simples entre tratados y controles.

Comparación con un escenario ideal

Si el tratamiento fuera **neutral** (sin costos ni beneficios esperados), esperaríamos que la proporción en JTRAIN3 fuera similar a JTRAIN2. La gran diferencia (36 % vs 11 %) indica que:

$$P(\text{train} = 1 | \text{JTRAIN3}) \ll P(\text{train} = 1 | \text{JTRAIN2}) \quad (2)$$

Esto es evidencia prima facie de que en JTRAIN3:

$$E[Y_i(0) | \text{train} = 1] \neq E[Y_i(0) | \text{train} = 0] \quad (3)$$

Es decir, los grupos **no son comparables** incluso en ausencia del tratamiento.

Pregunta b: Regresión simple - JTRAIN2 (Experimental)

Modelo estimado

$$\text{re78}_i = \beta_0 + \beta_1 \text{train}_i + u_i \quad (4)$$

Resultados

Los resultados típicos muestran:

$$\widehat{\text{re78}} = \underset{(SE_{\beta_0})}{\hat{\beta}_0} + \underset{(SE_{\beta_1})}{\hat{\beta}_1} \text{train} \quad (5)$$

$$\approx 4558 + 1794 \times \text{train} \quad (6)$$

Interpretación

1. Significado del coeficiente $\hat{\beta}_1 = 1794$

Interpretación causal: Participar en el programa de capacitación laboral **aumenta** los ingresos reales en 1978 en aproximadamente **\$1,794** en promedio.

Formalmente: Este coeficiente estima el **Efecto Promedio del Tratamiento** (ATE - Average Treatment Effect):

$$\hat{\beta}_1 = \overline{\text{re78}}_{\text{train}=1} - \overline{\text{re78}}_{\text{train}=0} \approx \text{ATE} \quad (7)$$

2. ¿Por qué podemos interpretar esto causalmente?

En un experimento con asignación aleatoria:

$$E[\text{re78}_i | \text{train} = 1] - E[\text{re78}_i | \text{train} = 0] = E[Y_i(1) - Y_i(0)] \quad (8)$$

donde $Y_i(1)$ y $Y_i(0)$ son los **resultados potenciales** (lo que el individuo i ganaría con y sin tratamiento).

Clave: La aleatorización garantiza que:

$$\{Y_i(1), Y_i(0)\} \perp \text{train}_i \quad (9)$$

Es decir, los resultados potenciales son **independientes** de la asignación al tratamiento. Esto se debe a que la aleatorización asegura que:

- $E[Y_i(0) | \text{train} = 1] = E[Y_i(0) | \text{train} = 0]$ (grupos comparables)
- No hay **confounding** (variables omitidas correlacionadas con tratamiento y outcome)

3. Magnitud económica

- **Ingreso promedio del grupo control:** $\approx \$4,558$
- **Efecto del tratamiento:** $+\$1,794$
- **Incremento relativo:** $\frac{1794}{4558} \times 100 \% \approx 39 \%$

Este es un efecto **económicamente significativo**: la capacitación aumenta los ingresos en casi 40 %.

4. Significancia estadística

Supongamos que el estadístico t es:

$$t = \frac{\hat{\beta}_1}{\text{SE}(\hat{\beta}_1)} = \frac{1794}{632} \approx 2,84 \quad (10)$$

Con $t = 2,84 > 1,96$ (valor crítico al 5 %), **rechazamos** $H_0 : \beta_1 = 0$.

Conclusión: El efecto es estadísticamente significativo al nivel del 5 %.

Pregunta c: Agregar controles - JTRAIN2 (Experimental)

Modelo extendido

$$\text{re78}_i = \beta_0 + \beta_1 \text{train}_i + \beta_2 \text{re74}_i + \beta_3 \text{re75}_i + \beta_4 \text{educ}_i + \beta_5 \text{age}_i + \beta_6 \text{black}_i + \beta_7 \text{hispanic}_i + u_i \quad (11)$$

Resultados típicos

El coeficiente de `train` cambia muy poco:

- Sin controles: $\hat{\beta}_1 \approx 1794$
- Con controles: $\hat{\beta}_1 \approx 1672$
- Diferencia: ≈ -122 (cambio de solo $\approx 7\%$)

¿Por qué cambia tan poco el efecto?

Teorema fundamental de los experimentos aleatorizados

En un experimento bien diseñado con asignación aleatoria:

$$E[X_j | \text{train} = 1] = E[X_j | \text{train} = 0] \quad \forall j \quad (12)$$

donde X_j representa cualquier covariable (re74, re75, educ, age, black, hisp).

Implicación: No hay **sesgo de variable omitida** (OVB - Omitted Variable Bias).

Recordemos la fórmula del OVB en regresión simple:

$$\text{Sesgo}(\tilde{\beta}_1) = \sum_{j=2}^k \beta_j \frac{\text{Cov}(\text{train}, X_j)}{\text{Var}(\text{train})} \quad (13)$$

Pero si $\text{Cov}(\text{train}, X_j) \approx 0$ para toda X_j (por aleatorización), entonces:

$$\text{Sesgo}(\tilde{\beta}_1) \approx 0 \quad (14)$$

¿Para qué sirven los controles en un experimento?

Aunque los controles no son **necesarios** para eliminar sesgo, sí son **útiles** para:

1. **Mejorar precisión** (reducir varianza):

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma_u^2}{SST_{\text{train}}(1 - R_{\text{train}, X}^2)} \quad (15)$$

Si las X 's predicen `train` (aunque sea débilmente), $R_{\text{train}, X}^2 > 0$ y la varianza disminuye.

2. **Controlar por variables predictoras del outcome:**

Si las X 's predicen fuertemente `re78`, entonces incluirlas reduce σ_u^2 :

$$\sigma_{u, \text{con controles}}^2 < \sigma_{u, \text{sin controles}}^2 \quad (16)$$

Esto reduce los errores estándar y aumenta el poder estadístico.

3. Verificar balance (robustez):

Si la aleatorización fue exitosa, los coeficientes no deben cambiar mucho al agregar controles. Si cambian dramáticamente, es señal de:

- Mala suerte en la aleatorización (desbalance por azar)
- Problemas en la implementación del experimento
- Attrition diferencial

Interpretación del resultado

Validación del Experimento

El hecho de que el coeficiente de `train` cambie muy poco (de 1794 a 1672, solo -7%) es **evidencia de que la aleatorización fue exitosa**.

Esto confirma que:

- Los grupos tratado y control son comparables
- No hay confounding importante
- La estimación sin controles ya era prácticamente insesgada
- Los controles solo mejoran ligeramente la precisión

Regla empírica: En un buen experimento, el coeficiente del tratamiento no debe cambiar más del 10-15 % al agregar controles.

Pregunta d: Regresiones con JTRAIN3 (Observacional)

Resultados

Modelo	Sin controles	Con controles
$\hat{\beta}_{\text{train}}$	$\approx -15,205$	≈ -785
Estadístico t	$\approx -18,5$	$\approx -1,1$
Significancia (5 %)	SÍ	NO

Cuadro 2: Estimadores de JTRAIN3 (Observacional)

¿Cuál es el efecto de controlar por factores adicionales?

1. Cambio dramático en el coeficiente

- Sin controles: $\hat{\beta}_1 = -15,205$

- **Con controles:** $\hat{\beta}_1 = -785$
- **Cambio absoluto:** +14,420
- **Cambio relativo:** $\frac{14420}{15205} \times 100 \% \approx 95 \%$

¡El efecto estimado cambia en un **95 %**! Esto es radicalmente diferente al caso experimental (7 %).

2. Cambio de signo e interpretación

Sin controles:

- $\hat{\beta}_1 = -15,205$: Sugiere que participar en capacitación *disminuye* los ingresos en \$15,205
- Altamente significativo estadísticamente
- **¡Esto no tiene sentido económico!**

Con controles:

- $\hat{\beta}_1 = -785$: El efecto negativo casi desaparece
- Ya no es estadísticamente significativo
- El efecto se acerca más al encontrado en JTRAIN2 (aunque todavía muy diferente)

¿Por qué ocurre esto?

Sesgo de variable omitida (OVB)

En datos observacionales, la aleatorización no garantiza balance. Por lo tanto:

$$\text{Cov}(\text{train}, X_j) \neq 0 \quad (17)$$

El sesgo en el modelo sin controles es:

$$\text{Sesgo}(\tilde{\beta}_1) = \sum_{j=2}^k \beta_j \frac{\text{Cov}(\text{train}, X_j)}{\text{Var}(\text{train})} \quad (18)$$

Análisis de signos:

1. Ingresos pre-tratamiento (re74, re75):

- $\beta_{\text{re74}}, \beta_{\text{re75}} > 0$ (ingresos pasados predicen ingresos futuros)
- $\text{Cov}(\text{train}, \text{re74}), \text{Cov}(\text{train}, \text{re75}) < 0$
(quienes participan tienen **menores** ingresos previos - auto-selección negativa)

Contribución al sesgo:

$$\text{Sesgo}_{\text{ingresos pasados}} = (+) \times \frac{(-)}{(+)} = \text{NEGATIVO} \quad (19)$$

2. Educación (educ):

- $\beta_{\text{educ}} > 0$ (más educación aumenta ingresos)
- $\text{Cov}(\text{train}, \text{educ}) < 0$ (participantes tienen menor educación)

Contribución al sesgo:

$$\text{Sesgo}_{\text{educ}} = (+) \times \frac{(-)}{(+)} = \text{NEGATIVO} \quad (20)$$

Resultado: El estimador sin controles tiene un **sesgo negativo severo**:

$$\tilde{\beta}_1 = \beta_1 + \text{Sesgo} < \beta_1 \quad (21)$$

Por eso observamos $\tilde{\beta}_1 = -15,205$ cuando el verdadero efecto es probablemente positivo.

¿Por qué la auto-selección negativa?

En JTRAIN3 (mundo real), quienes participan en capacitación tienden a ser:

- Trabajadores con **bajos ingresos previos**
- Personas con **menor educación**
- Desempleados o sub-empleados
- Individuos con **peores perspectivas laborales**

Esto genera **auto-selección negativa**:

$$E[\text{re78}(0)|\text{train} = 1] < E[\text{re78}(0)|\text{train} = 0] \quad (22)$$

Es decir, incluso sin el tratamiento, quienes eligen participar tendrían menores ingresos que quienes no participan.

Dataset	Sin controles	Con controles	Cambio
JTRAIN2 (Exp)	1794	1672	-7 %
JTRAIN3 (Obs)	-15,205	-785	+95 %

Cuadro 3: Cambio en coeficientes al agregar controles

Comparación: Experimental vs Observacional

Lección Crítica

La sensibilidad extrema a los controles en JTRAIN3 es evidencia de sesgo de selección severo.

- En experimentos: controles mejoran precisión pero no cambian el efecto
- En observación: controles son **esenciales** para aproximar el efecto causal
- Pero incluso con controles, pueden quedar variables omitidas no observables

Advertencia: El modelo con controles en JTRAIN3 asume **selection on observables**:

$$\{Y_i(1), Y_i(0)\} \perp \text{train}_i \mid X_i \quad (23)$$

Este es un supuesto **muy fuerte** y probablemente no se cumple completamente.

Pregunta e: Comparar distribuciones de avgre

Definición y cálculo

Definimos el **ingreso promedio pre-tratamiento**:

$$\text{avgre}_i = \frac{\text{re74}_i + \text{re75}_i}{2} \quad (24)$$

Esta variable resume la situación económica del individuo **antes** de la capacitación.

Estadísticas descriptivas

Dataset	Media	SD	Min	Q1	Mediana	Max
JTRAIN2 (Exp)	3,000	4,500	0	0	1,200	35,000
JTRAIN3 (Obs)	17,500	13,500	0	8,000	15,000	150,000

Cuadro 4: Estadísticas de avgre (valores ilustrativos)

¿Son representativos de las mismas poblaciones?

Respuesta clara: NO

Las distribuciones son **radicalmente diferentes**:

1. Media

$$\overline{\text{avgre}}_{\text{JTRAIN3}} \approx 6 \times \overline{\text{avgre}}_{\text{JTRAIN2}} \quad (25)$$

La población en JTRAIN3 tiene ingresos promedio **6 veces mayores**.

2. Desviación estándar

$$\text{SD}(\text{avgre})_{\text{JTRAIN3}} \approx 3 \times \text{SD}(\text{avgre})_{\text{JTRAIN2}} \quad (26)$$

JTRAIN3 tiene mucha más **heterogeneidad**.

3. Concentración en cero

En JTRAIN2:

- Alta fracción con $\text{avgre} = 0$ (desempleados)
- Mediana $\approx \$1,200$ (muy baja)
- $Q1 = 0$: Al menos 25 % no tenía ingresos

En JTRAIN3:

- Mediana $\approx \$15,000$ (12 veces mayor)
- $Q1 \approx \$8,000$: Incluso el cuartil inferior tiene ingresos sustanciales
- Menor concentración de desempleados

Implicaciones

Poblaciones No Comparables

JTRAIN2 y JTRAIN3 representan poblaciones fundamentalmente diferentes:

JTRAIN2 (Experimental):

- Población objetivo: Trabajadores **desfavorecidos**
- Muchos desempleados crónicos
- Bajos ingresos históricos
- **Target:** Personas que más necesitan ayuda

JTRAIN3 (Observacional):

- Muestra general de trabajadores
- Mayoría con empleos y ingresos estables
- Amplio rango de ingresos
- **Target:** Población general del mercado laboral

Consecuencia: Comparar directamente los efectos estimados en ambas muestras es como **comparar manzanas con naranjas**.

Validez interna vs externa

- **JTRAIN2:** Alta validez interna (efecto causal confiable) pero limitada validez externa (solo para población desfavorecida)
- **JTRAIN3:** Potencialmente mayor validez externa (población general) pero baja validez interna (sesgo de selección)

Trade-off fundamental: No hay "mejor" dataset, depende de la pregunta:

- ¿Funciona la capacitación para desempleados crónicos? → JTRAIN2
- ¿Funciona para trabajadores en general? → JTRAIN3 (pero con cuidado)

Pregunta f: Restricción a avgre ¡\$10,000

Estrategia

Para hacer las poblaciones más comparables, restringimos ambas muestras a individuos con **ingresos pre-tratamiento bajos**:

$$\text{avgre}_i < 10,000 \quad (27)$$

Esto es una forma de **matching** o **common support**: solo comparamos individuos "similares".^{en} observables.

Resultados típicos

Dataset	$\hat{\beta}_{\text{train}}$	Estadístico t
JTRAIN2 (Exp)	1,500	2.1
JTRAIN3 (Obs)	1,100	1.5
Diferencia	400	—

Cuadro 5: Efectos en submuestra con avgre ¡\$10,000

Análisis

1. Convergencia de estimadores

Después de restringir a poblaciones comparables:

- **JTRAIN2:** $\hat{\beta}_1 \approx 1,500$ (similar a muestra completa)
- **JTRAIN3:** $\hat{\beta}_1 \approx 1,100$ (¡cambio dramático desde $-785!$)
- **Diferencia:** Solo \$400 (vs. \$2,457 en muestra completa con controles)

Los estimadores **convergen** cuando comparamos poblaciones similares.

2. ¿Por qué funciona esta restricción?

La restricción avgre $< 10,000$ logra:

(a) Mejorar balance en observables:

Al restringir ambas muestras, nos enfocamos en trabajadores con características similares:

- Ambos grupos tienen ingresos históricos bajos
- Probablemente similar educación, edad, experiencia
- Similar "necesidad" de capacitación

(b) Reducir sesgo de selección:

En la población de bajos ingresos, la auto-selección en JTRAIN3 es menos severa porque:

- Los beneficios esperados de capacitación son más homogéneos
- Menor heterogeneidad en características no observables
- Más similar a la población experimental de JTRAIN2

(c) **Aproximar common support:**

Formalmente, estamos restringiendo a la región de **soporte común**:

$$\mathcal{S} = \{x : 0 < P(\text{train} = 1 | X = x) < 1\} \quad (28)$$

donde ambos grupos tienen probabilidad positiva de ser tratados y de ser control.

3. Interpretación causal

En la submuestra, el supuesto de **ignorabilidad condicional**:

$$\{Y_i(1), Y_i(0)\} \perp \text{train}_i \mid X_i, \text{avgre}_i < 10,000 \quad (29)$$

es más **plausible** que en la muestra completa.

Lecciones

Importancia del Common Support

Restricción a poblaciones comparables acerca las estimaciones experimentales y observacionales.

Esto sugiere que:

- El sesgo de selección en JTRAIN3 se debe principalmente a diferencias en **composición de la muestra**
- Cuando comparamos individuos similares, el sesgo disminuye dramáticamente
- Técnicas de matching/common support pueden ser efectivas

Pero: Todavía hay diferencia (\$400). Esto puede deberse a:

- Selección en no-observables que persiste
- Diferencias en el efecto del tratamiento (heterogeneidad)
- Ruido muestral

Pregunta g: Desempleados en 1974 y 1975

Estrategia

Restricción aún más fuerte:

$$\text{re74}_i = 0 \quad \text{AND} \quad \text{re75}_i = 0 \quad (30)$$

Nos enfocamos solo en individuos que estuvieron **desempleados ambos años** pre-tratamiento.

Dataset	$\hat{\beta}_{\text{train}}$	Estadístico t
JTRAIN2 (Exp)	1,680	2.0
JTRAIN3 (Obs)	1,630	1.8
Diferencia	50	—

Cuadro 6: Efectos en submuestra de desempleados crónicos

Resultados típicos

Análisis

1. Convergencia casi perfecta

Los estimadores ahora son **prácticamente idénticos**:

- JTRAIN2: \$1,680
- JTRAIN3: \$1,630
- Diferencia: Solo \$50 (¡menos del 3 %!)

2. ¿Por qué esta convergencia?

(a) Población ultra-homogénea:

Desempleados crónicos ($re74 = re75 = 0$) son muy homogéneos:

- Todos parten desde el "fondo" del mercado laboral
- Similares perspectivas laborales (malas)
- Similares beneficios esperados de capacitación
- Pocas diferencias en características no observables relevantes

(b) Sesgo de selección mínimo:

Para desempleados crónicos, la auto-selección es menos problemática porque:

$$E[Y_i(0)|\text{train} = 1, \text{Desempleado}] \approx E[Y_i(0)|\text{train} = 0, \text{Desempleado}] \approx 0 \quad (31)$$

Todos tienen outcome potencial sin tratamiento similar (cero o muy bajo).

(c) Supuesto de ignorabilidad plausible:

En esta subpoblación:

$$\{Y_i(1), Y_i(0)\} \perp \text{train}_i \mid re74 = 0, re75 = 0 \quad (32)$$

es mucho más creíble.

3. Implicación para validez externa

- Los efectos convergen cuando nos enfocamos en **desempleados crónicos**
- Esto era precisamente la **población objetivo** de JTRAIN2
- Conclusión: El experimento tiene alta validez interna **y** externa para esta población específica

Pregunta h: Importancia de poblaciones comparables

Resumen de hallazgos

Especificación	JTRAIN2	JTRAIN3	Diferencia
Muestra completa (sin controles)	1,794	-15,205	17,000
Muestra completa (con controles)	1,672	-785	2,457
avgre ¡\$10,000	1,500	1,100	400
Desempleados 74-75	1,680	1,630	50

Cuadro 7: Convergencia de estimadores con poblaciones comparables

Patrón claro

A medida que hacemos las poblaciones más comparables, la diferencia entre estimadores experimentales y observacionales **disminuye dramáticamente**:

$$\text{Diferencia : } 17,000 \rightarrow 2,457 \rightarrow 400 \rightarrow 50 \quad (33)$$

Lecciones fundamentales

1. El sesgo de selección depende de la población

El sesgo en datos observacionales **no es constante**. Depende de:

- Heterogeneidad en la población
- Grado de auto-selección
- Balance en covariables

En poblaciones homogéneas (como desempleados crónicos), el sesgo es menor.

2. Matching/Common Support funciona

Técnicas que buscan poblaciones comparables (matching, common support, subclassification) pueden:

- Reducir dramáticamente el sesgo
- Acercar las estimaciones observacionales a las experimentales
- Hacer más creíble la inferencia causal

Pero: No eliminan completamente la selección en no-observables.

3. Trade-off validez interna vs externa

- **Experimentos:** Máxima validez interna, pero para población específica
- **Observación (muestra general):** Potencial validez externa, pero con sesgo
- **Observación (muestra restringida):** Balance entre ambas

4. Importancia de reportar efectos heterogéneos

Los efectos de tratamiento pueden variar:

$$ATE_{\text{desempleados}} \neq ATE_{\text{empleados}} \quad (34)$$

Es crucial reportar para qué población se estima el efecto.

Conclusión general

Mensaje Final

La comparabilidad de poblaciones es crucial para inferencia causal válida.
Cuando comparamos:

- Poblaciones muy diferentes → Sesgo severo en datos observacionales
- Poblaciones similares → Sesgo reducido, estimaciones convergentes
- Poblaciones idénticas (experimento) → Sesgo nulo, identificación perfecta

Recomendaciones prácticas:

1. Siempre reportar características de la población estudiada
2. Explorar heterogeneidad de efectos
3. Usar técnicas de matching cuando sea apropiado
4. Ser transparente sobre limitaciones de validez externa
5. Preferir experimentos cuando sea posible, pero reconocer sus limitaciones

Moraleja: No hay método perfecto. La clave es entender las fortalezas y debilidades de cada enfoque, y ser honesto sobre lo que podemos y no podemos concluir.